



ENTERED

1600

## RAW SEQUENCE LISTING

DATE: 09/12/2002

PATENT APPLICATION: US/10/042,059B

TIME: 14:58:41

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

3 <110> APPLICANT: Gellissen, Gerd  
 4 Braus, Gerhard  
 5 Pries, Ralph  
 6 Krappmann, Sven  
 7 Strasser, Alexander  
 9 <120> TITLE OF INVENTION: Nucleic Acid Molecule Comprising a Nucleic Acid Coding for a  
 Polypeptide  
 10 with Chorismate Mutase Activity  
 12 <130> FILE REFERENCE: 029474-5007-00  
 14 <140> CURRENT APPLICATION NUMBER: 10/042059B  
 15 <141> CURRENT FILING DATE: 2001-10-25  
 17 <150> PRIOR APPLICATION NUMBER: DE 199 19 124.7  
 18 <151> PRIOR FILING DATE: 1999-04-27  
 20 <160> NUMBER OF SEQ ID NOS: 7  
 22 <170> SOFTWARE: PatentIn version 3.1  
 24 <210> SEQ ID NO: 1  
 25 <211> LENGTH: 843  
 26 <212> TYPE: DNA  
 27 <213> ORGANISM: Hansenula polymorpha  
 29 <220> FEATURE:  
 30 <221> NAME/KEY: CDS  
 31 <222> LOCATION: (1)..(843)  
 32 <223> OTHER INFORMATION:  
 34 <400> SEQUENCE: 1  
 35 atg gac ttt atg aag cca gaa aca gtg ctg gac ctt ggc aac att aga 48  
 36 Met Asp Phe Met Lys Pro Glu Thr Val Leu Asp Leu Gly Asn Ile Arg  
 37 1 5 10 15  
 39 gat gcc ttg gtc cgg atg gag gat acg atc atc ttc aac ttt atc gag 96  
 40 Asp Ala Leu Val Arg Met Glu Asp Thr Ile Ile Phe Asn Phe Ile Glu  
 41 20 25 30  
 43 cgg tcg cag ttc tat gcg tcg ccc tcg gta tac aaa gtc aac cag ttc 144  
 44 Arg Ser Gln Phe Tyr Ala Ser Pro Ser Val Tyr Lys Val Asn Gln Phe  
 45 35 40 45  
 47 cct att ccc aac ttc gac ggc tcg ttc ttg gac tgg ctg ttg tcg cag 192  
 48 Pro Ile Pro Asn Phe Asp Gly Ser Phe Leu Asp Trp Leu Leu Ser Gln  
 49 50 55 60  
 51 cac gag cga atc cat tcg cag gtg agg aga tac gac gcg cca gac gag 240  
 52 His Glu Arg Ile His Ser Gln Val Arg Arg Tyr Asp Ala Pro Asp Glu  
 53 65 70 75 80  
 55 gtg cct ttt ttc ccc aac gtg ctg gaa aaa acg ttt ctg ccc aag atc 288  
 56 Val Pro Phe Phe Pro Asn Val Leu Glu Lys Thr Phe Leu Pro Lys Ile  
 57 85 90 95  
 59 aac tac cca tcg gtg cta gcc tcc tac gcg gat gaa atc aac gtc aac 336  
 60 Asn Tyr Pro Ser Val Leu Ala Ser Tyr Ala Asp Glu Ile Asn Val Asn

## RAW SEQUENCE LISTING

DATE: 09/12/2002

PATENT APPLICATION: US/10/042,059B

TIME: 14:58:41

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

```

61          100          105          110
63 aaa gag ata ctc aag atc tac acg tca gag ata gta cca gga ata gct      384
64 Lys Glu Ile Leu Lys Ile Tyr Thr Ser Glu Ile Val Pro Gly Ile Ala
65          115          120          125
67 gca ggc agc gga gag cag gag gac aac ctt ggc tcg tgc gca atg gcc      432
68 Ala Gly Ser Gly Glu Gln Glu Asp Asn Leu Gly Ser Cys Ala Met Ala
69          130          135          140
71 gac atc gag tgc ctg cag tcg cta tcc aga aga atc cat ttt ggc cgt      480
72 Asp Ile Glu Cys Leu Gln Ser Leu Ser Arg Arg Ile His Phe Gly Arg
73 145          150          155          160
75 ttt gtc gca gag gct aaa ttt atc agt gag ggg gac aag att gtg gat      528
76 Phe Val Ala Glu Ala Lys Phe Ile Ser Glu Gly Asp Lys Ile Val Asp
77          165          170          175
79 ctg atc aaa aag aga gat gtg gaa ggc att gag gcg ctc atc aca aac      576
80 Leu Ile Lys Lys Arg Asp Val Glu Gly Ile Glu Ala Leu Ile Thr Asn
81          180          185          190
83 gcc gag gtc gaa aaa cgg atc ttg gac aga ctt ctg gag aag gga agg      624
84 Ala Glu Val Glu Lys Arg Ile Leu Asp Arg Leu Leu Glu Lys Gly Arg
85          195          200          205
87 gcg tat gga aca gac ccg aca cta aag ttc acg cag cac att cag agc      672
88 Ala Tyr Gly Thr Asp Pro Thr Leu Lys Phe Thr Gln His Ile Gln Ser
89          210          215          220
91 aag gtg aag ccc gag gtg att gtg aaa atc tac aag gat ttc gtg att      720
92 Lys Val Lys Pro Glu Val Ile Val Lys Ile Tyr Lys Asp Phe Val Ile
93 225          230          235          240
95 ccg ctc acg aag aag gtc gaa gtc gac tac ttg ctg aga cgg ctg gag      768
96 Pro Leu Thr Lys Lys Val Glu Val Asp Tyr Leu Leu Arg Arg Leu Glu
97          245          250          255
99 gac gag gag gac gat gat gcg acg cag aaa agc ggc ggc tac gtt gac      816
100 Asp Glu Glu Asp Asp Asp Ala Thr Gln Lys Ser Gly Gly Tyr Val Asp
101          260          265          270
103 cgg ttt ctc tcc tct ggc ttg tac tag      843
104 Arg Phe Leu Ser Ser Gly Leu Tyr
105          275          280
108 <210> SEQ ID NO: 2
109 <211> LENGTH: 280
110 <212> TYPE: PRT
111 <213> ORGANISM: Hansenula polymorpha
113 <400> SEQUENCE: 2.
115 Met Asp Phe Met Lys Pro Glu Thr Val Leu Asp Leu Gly Asn Ile Arg
116 1          5          10          15
119 Asp Ala Leu Val Arg Met Glu Asp Thr Ile Ile Phe Asn Phe Ile Glu
120          20          25          30
123 Arg Ser Gln Phe Tyr Ala Ser Pro Ser Val Tyr Lys Val Asn Gln Phe
124          35          40          45
127 Pro Ile Pro Asn Phe Asp Gly Ser Phe Leu Asp Trp Leu Leu Ser Gln
128          50          55          60
131 His Glu Arg Ile His Ser Gln Val Arg Arg Tyr Asp Ala Pro Asp Glu
132 65          70          75          80

```

## RAW SEQUENCE LISTING

DATE: 09/12/2002

PATENT APPLICATION: US/10/042,059B

TIME: 14:58:41

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

```

135 Val Pro Phe Phe Pro Asn Val Leu Glu Lys Thr Phe Leu Pro Lys Ile
136                               85                               90                               95
139 Asn Tyr Pro Ser Val Leu Ala Ser Tyr Ala Asp Glu Ile Asn Val Asn
140                               100                              105                              110
143 Lys Glu Ile Leu Lys Ile Tyr Thr Ser Glu Ile Val Pro Gly Ile Ala
144                               115                              120                              125
147 Ala Gly Ser Gly Glu Gln Glu Asp Asn Leu Gly Ser Cys Ala Met Ala
148                               130                              135                              140
151 Asp Ile Glu Cys Leu Gln Ser Leu Ser Arg Arg Ile His Phe Gly Arg
152 145                               150                              155                              160
155 Phe Val Ala Glu Ala Lys Phe Ile Ser Glu Gly Asp Lys Ile Val Asp
156                               165                              170                              175
159 Leu Ile Lys Lys Arg Asp Val Glu Gly Ile Glu Ala Leu Ile Thr Asn
160                               180                              185                              190
163 Ala Glu Val Glu Lys Arg Ile Leu Asp Arg Leu Leu Glu Lys Gly Arg
164                               195                              200                              205
167 Ala Tyr Gly Thr Asp Pro Thr Leu Lys Phe Thr Gln His Ile Gln Ser
168                               210                              215                              220
171 Lys Val Lys Pro Glu Val Ile Val Lys Ile Tyr Lys Asp Phe Val Ile
172 225                               230                              235                              240
175 Pro Leu Thr Lys Lys Val Glu Val Asp Tyr Leu Leu Arg Arg Leu Glu
176                               245                              250                              255
179 Asp Glu Glu Asp Asp Asp Ala Thr Gln Lys Ser Gly Gly Tyr Val Asp
180                               260                              265                              270
183 Arg Phe Leu Ser Ser Gly Leu Tyr
184                               275                              280

```

187 &lt;210&gt; SEQ ID NO: 3

188 &lt;211&gt; LENGTH: 1655

189 &lt;212&gt; TYPE: DNA

190 &lt;213&gt; ORGANISM: Hansenula polymorpha

192 &lt;220&gt; FEATURE:

193 &lt;221&gt; NAME/KEY: gene

194 &lt;222&gt; LOCATION: (1)..(1655)

195 &lt;223&gt; OTHER INFORMATION: 1,8 kb genomic DNA-fragment from Hansenula polymorpha

198 &lt;220&gt; FEATURE:

199 &lt;221&gt; NAME/KEY: gene

200 &lt;222&gt; LOCATION: (1)..(1655)

201 &lt;223&gt; OTHER INFORMATION: 1,8 kb genomic DNA-fragment from Hansenula polymorpha

204 &lt;400&gt; SEQUENCE: 3

```

205 cccggcccaa tgccagcaat atggagacgt ttaggcagaa taggcgttcc atacttctca      60
207 cgctgcttgt tgccaccgga atatacaccg cattgcagtt tgcacacatc atactatatg      120
209 acgattacat tggcggaacg tatcgcgagt cgctcacgag acgcattaga atgacagaga      180
211 aatcgcgaaa cgaccttata gacgcacgtg aaaactacgg gtttggaggc agcaaggagg      240
213 agcgaatcca gcggtttttg tggttcagac atctttcgtg gcttttaggc gaggataagc      300
215 gaacttgagg agcgtttttt ttttcctgtt tagtttttgt aggtatggac tttatgaagc      360
217 cagaaacagt gctggacctt ggcaacatta gagatgcctt ggtccggatg gaggatacga      420
219 tcatcttcaa ctttatcgag cggtcgcagt tctatgcgtc gccctcggta taaaagtca      480
221 accagttccc tattcccaac ttgcacggct cgttcttgga ctggctgttg tcgcagcacg      540
223 agcgaatcca ttgcaggtg aggagatacg acgcgccaga cgaggtgcct ttttcccca      600

```

## RAW SEQUENCE LISTING

DATE: 09/12/2002

PATENT APPLICATION: US/10/042,059B

TIME: 14:58:41

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

```

225 acgtgctgga aaaaacgttt ctgcccaga tcaactaccc atcggtgcta gcctcctacg 660
227 cggatgaaat caacgtcaac aaagagatac tcaagatcta cagctcagag atagtaccag 720
229 gaatagctgc aggcagcgga gagcaggagg acaaccttgg ctctgctgca atggccgaca 780
231 tcgagtgcct gcagtcgcta tccagaagaa tccattttgg ccgttttgtc gcagaggcta 840
233 aatttatcag tgaggggggac aagatttgtg atctgatcaa aaagagagat gtggaaggca 900
235 ttgaggcgct catcacaac gccgaggtcg aaaaacggat cttggacaga cttctggaga 960
237 aggggaaggc gtatggaaca gaccgacac taaagtacac gcagcacatt cagagcaagg 1020
239 tgaagcccga ggtgattgtg aaaatctaca aggatttcgt gattccgctc acgaagaagg 1080
241 tcgaagtcga ctacttgctg agacggctgg aggacgagga ggacgatgat gcgacgcaga 1140
243 aaagcggcgg ctacgttgac cggtttctct cctctggctt gtactagaaa ttaaaatttt 1200
245 cagtacttta attattctcg aattctagtt cagataccgc atggtaattt caaaggccag 1260
247 aaaagtggcc gcgttggtcg gggcagctct cagaatagtc ggcgagaatc ctttgactag 1320
249 cccccaggca ccgctctgtc tccaaatacc cctaatagtc tcaacagcat ttctataaac 1380
251 cagcttcttg tagttgtccg tctgcatgtt ggacttgatc acatcgatcg gataaatact 1440
253 gaaccacatc ccgtaacctg ccagcgcccc aaagacgcag agcttccagt tctcgatgtc 1500
255 ctctctggca atattccgcg actcgatctc gtttttcacg agagcttcaa aagtcagaaa 1560
257 atacgctccg ctacccaaac tttctcttgc cagcgtaggt cccagacccc ggtagattaa 1620
259 cttgatgcct cccgtatggt acagcttctt gatcc 1655
262 <210> SEQ ID NO: 4
263 <211> LENGTH: 20
264 <212> TYPE: DNA
C--> 265 <213> ORGANISM: Artificial
267 <220> FEATURE:
268 <223> OTHER INFORMATION: Oligonucleotide
270 <400> SEQUENCE: 4
271 aattaaccct cactaaaggg 20
274 <210> SEQ ID NO: 5
275 <211> LENGTH: 22
276 <212> TYPE: DNA
C--> 277 <213> ORGANISM: Artificial
279 <220> FEATURE:
280 <223> OTHER INFORMATION: Oligonucleotide
282 <400> SEQUENCE: 5
283 gtaatacgac tcaactatagg gc 22
286 <210> SEQ ID NO: 6
287 <211> LENGTH: 26
288 <212> TYPE: DNA
C--> 289 <213> ORGANISM: Artificial
291 <220> FEATURE:
292 <223> OTHER INFORMATION: Oligonucleotide
294 <400> SEQUENCE: 6
295 atatagatct acaaaaacta aacagg 26
298 <210> SEQ ID NO: 7
299 <211> LENGTH: 28
300 <212> TYPE: DNA
C--> 301 <213> ORGANISM: Artificial
303 <220> FEATURE:
304 <223> OTHER INFORMATION: Oligonucleotide
306 <400> SEQUENCE: 7

```

RAW SEQUENCE LISTING

DATE: 09/12/2002

PATENT APPLICATION: US/10/042,059B

TIME: 14:58:41

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

307 atatagatct gatgcgacgc agaaaagc

28

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/042,059B

DATE: 09/12/2002  
TIME: 14:58:42

Input Set : A:\EP.txt  
Output Set: N:\CRF4\09122002\J042059B.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 9

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,  
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:4,5,6,7

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/042,059B

DATE: 09/12/2002

TIME: 14:58:42

Input Set : A:\EP.txt

Output Set: N:\CRF4\09122002\J042059B.raw

L:265 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:4  
L:277 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:5  
L:289 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:6  
L:301 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:7